

1 Cultural conformity generates extremely stable traditions 2 in bird song

3 Robert F. Lachlan^{1*}, Oliver Ratmann² and Stephen Nowicki³

4 **1** Department of Biological and Experimental Psychology, Queen Mary University of London,
5 London, E1 4NS, United Kingdom;

6 **2** Department of Mathematics, Imperial College London, London, SW7 2AZ, United Kingdom;

7 **3** Department of Biology, Duke University, Durham, NC 27708 United States and Department
8 of Neurobiology, Duke University Medical School, Durham, NC 27708.

9

10

11 Abstract

12 Cultural traditions have been observed in a wide variety of animal species. It remains unclear,
13 however, what is required for social learning to give rise to stable traditions: what level of
14 precision and what learning strategies are required. We address these questions by fitting models
15 of cultural evolution to learned bird song. We recorded 615 swamp sparrow (*Melospiza*
16 *georgiana*) song repertoires, and compared syllable frequency distributions to the output of
17 individual-based simulations. We find that syllables are learned with an estimated error rate of
18 1.85% and with a conformist bias in learning. This bias is consistent with a simple mechanism of
19 overproduction and selective attrition. Finally, we estimate that syllable types could frequently
20 persist for more than 500 years. Our results demonstrate conformist bias in natural animal
21 behaviour and show that this, along with moderately precise learning, may support traditions
22 whose stability rivals those of humans.

23 Social learning underlies a wide range of behaviour in many animal species. Yet while cultural
24 transmission is widespread among animals, human culture is unique in its complexity and
25 tendency to accumulate adaptive variations^{1,2}. One proposed explanation for this apparent
26 paradox is that cultural transmission in animals relies on mechanisms that are simply not precise
27 enough to support long-lasting traditions shared by large numbers of individuals, and that this
28 prevents the accumulation of complex culture^{1,3-6}. It has also been argued that humans are unique
29 in the types of strategies we employ to decide who to learn from, and these strategies may help
30 us maintain stable traditions in the face of innovation and errors in transmission⁷⁻¹⁰. Two central
31 challenges, therefore, are to understand the social learning processes underlying animal
32 traditions, and to assess how stable such traditions may be.

33 Of various social learning strategies, conformity has been most studied^{7,10}. If individuals sample
34 the behavioural variants of a subset of their population, and select variants to learn from at
35 random from within that subset, then they will select variants with probabilities approximating
36 the frequency of those variants in the population. If individuals have a conformist bias, they are
37 more likely to pick a common variant from the sampled set than expected by chance⁷.

38 Conformist biases therefore cause common variants in the population to increase in frequency
39 and select against rare variants. Because new variants are rare, conformist biases tend to remove
40 them, theoretically allowing stable traditions even if innovation rates are high. Often regarded as
41 central to human culture¹⁰, the evidence for conformist biases in animals is growing. This
42 evidence includes experimental and observational evidence that demonstrates animals are paying
43 attention to group norms, and the behaviour of the majority¹¹⁻¹⁶. Explicit tests of conformist
44 biases (that is a disproportionate tendency to copy the majority) are rarer, and include only one

study in captive sticklebacks¹⁷, and two studies of free-living population of great tits interacting with experimental equipment^{18,19}. Together, these studies suggest that animals may frequently employ conformist biases to decide whom to learn from. In this study, we examine whether conformist biases are present in an example of naturally occurring behaviour. Specifically, we focus on learned song in six populations of swamp sparrows in the Eastern United States. Using computational models, we assess the implications of conformist bias on the stability of these cultural traits.

Song in oscine songbirds is a well-studied example of animal social learning and cultural evolution^{20,21}. To produce species-typical songs, young birds must memorize songs heard early in life, and later develop imitations of those songs. Evidence from hand-raised birds of several species, including swamp sparrows, demonstrates that individuals are capable of imitating model songs with great precision²²⁻²⁴. Laboratory studies also have revealed details of the process of cultural transmission²⁵. Male swamp sparrows first memorize a set of syllable types (Fig. 1) during a sensitive period lasting the first 8 weeks of life²⁴. The following spring, they themselves begin to sing and develop precise renditions of around 3 (range 1-6) syllable types. A key feature of this process is that individuals initially produce approximately 13 syllable types (around half of which are similar enough to be regarded as imitations or modifications of demonstrator syllable types), and this is gradually reduced to their adult repertoire (85% of which, in a sample of hand-raised birds, were similar to tape-recorded demonstrator syllable types)^{22,23}. This process of overproduction and selective attrition has also been found in other species as well²⁵⁻²⁸, and provides an extended period during which individuals can select which songs to produce based on those they hear others singing. This process makes song learning a particularly interesting

case in which to examine conformist bias, which requires individuals to sample and compare more traits than they eventually produce²⁹.

Complementing laboratory experiments, field studies have demonstrated in some species of songbird that social learning might generate long-lasting traditions³⁰⁻³³. Many individuals within local populations may share the same song type³⁴, which suggests that learning is often precise^{35,36}. Direct evidence for stable traditions comes from diachronic studies showing that song types might persist for decades³⁰⁻³², but such studies are limited by the relatively recent advent of recording equipment. A recent analysis of songs in an isolated introduced population suggests that yellowhammer (*Emberiza citronella*) song types might persist for 100 years or more³³. Neither laboratory nor field studies have provided detailed quantitative estimates of learning precision, learning strategies, or the longevity of song traditions, however, due to the challenging of inherently small sample sizes associated with such studies.

Here we use Approximate Bayesian Computation (ABC)^{37,38} to overcome the latter problem, by fitting individual-based simulation models of cultural evolution to frequency distributions of syllable types from field recordings of 6 populations of swamp sparrows. ABC allows the estimation of posterior distributions of parameters for models that are too complex for likelihood calculations, but from which simulated data can readily be generated. It is thus particularly suited for studies of cultural transmission³⁹⁻⁴¹. By pairing ABC with individual-based simulations³⁸, we are able to fit complex models in which individuals use multiple learning strategies.

We assessed the strength of evidence for three fundamental categories of social learning biases: conformist biases (described above), demonstrator biases, and content biases^{8,42}. Demonstrator

biases (sometimes referred to as “model” biases) occur when individuals prefer to learn from particular demonstrators (often referred to as “tutors” in the song learning literature), irrespective of the trait they display. In the context of song learning, demonstrator biases might arise because some males sing more frequently than others, but also, hypothetically, if individuals prefer to learn from more successful males holding better territories. Content biases might occur when some variants of traits are inherently more attractive to learn than others regardless of who demonstrates them. In swamp sparrow song, previous studies have found no evidence that particular syllable types are inherently more attractive than others (see Supplementary Note 1), although some rare syllable types that we sampled appeared to deviate more from species-specific norms, which might suggest they are less likely to be copied, leading to content bias.

Each of these biases are expected to leave distinct imprints on observable frequency distributions of bird song syllables in populations²⁹ (Fig. 2a). Demonstrator biases reduce the effective population size, and thus reduce cultural diversity in a neutral model of evolution (much as bottlenecks reduce genetic diversity within populations), and hence increase the relative frequency of the commonest variants, and decrease the frequency of rare variants. Content biases result in many novel types in the population being ranked as unattractive, and being quickly removed from the population, and hence reduce the frequency of rare types in the population. Conformist biases increase the frequency of common variants. But if an individual can only sample a small proportion of the population, it is likely to sample both intermediate and rare variants only once, and thus rare variants are not disadvantaged much more than intermediate ones. All three biases might exist simultaneously, and thus our simulations took all three into account. But based on these observations (also see Fig. 2a), we expected slight differences in the

consequences of the different biases, and thus to be able to differentiate between them and determine the strength of evidence for each.

Results

Syllable diversity and geographic variation

We recorded the complete song repertoires of 615 adult male swamp sparrows from 6 different populations across the subspecies range of *M. g. georgiana* (see Supplementary Table 1 for details of the populations and samples). In these populations, swamp sparrows lived at high densities (typically $>150 \text{ km}^{-2}$). We then selected exemplars of each syllable type in each male's repertoire, based on a subjective assessment of recording quality, compared each of these syllables with every other syllable in the data-set using the implementation of dynamic time warping in Luscinia (<https://github.com/rflachlan/Luscinia>), hierarchically clustered syllables, and used the Global Silhouette Index⁴³ to determine the number of types. This analysis found a total of 160 different syllable types in our sample (Supplementary Fig. 2). Within populations, there was no evidence of syllables being preferentially shared with neighbours (Mantel test of correlation geographic distances between individuals and Jaccard index of repertoire sharing, using populations as strata, $r=-0.0044$, $p=0.28$), but syllable sharing was much higher within populations than between populations (Fig. 3), and syllable sharing over the whole data-set was clearly correlated with the distance between individuals (Mantel test of entire data-set, $r=-0.307$, $p<0.0002$). These results suggest that there is little or no geographic structure to demonstrator choices within a population, but that cultural divergence arises between populations. The frequency distribution of syllable types across the six population showed a long-tail distribution,

with some very common types and many very rare types in each population. The distributions were markedly similar between populations (Fig. 2b).

Cultural Evolutionary Models of Song Learning

We next constructed an individual-based simulation model of syllable learning. Each run of the model lasts for a period of 5000 “years”, and at the end of this period we measure the diversity of syllable types present in each population. In the model, there are a finite number, N_s , of possible syllable types, which are learned by individuals during the first year of life from adults that they selected from a population of N_I individuals. Mutations occur at a fixed rate, μ , per syllable-learning event, and generate a syllable type other than that produced by the demonstrator. “Mutations” encompassed innovations and errors, but also immigration of syllable types into the population. The mutated syllable type could already exist in the population, allowing for the possibility that individuals could re-invent syllables. Demonstrator biases in learning were incorporated by allocating each individual a value determining how attractive it would be as a demonstrator: the greater the variance, v , of this value, the greater the role of demonstrator bias. Content biases were incorporated by setting only a proportion of potential syllable types, p_{att} , as being attractive to learn. The smaller the value of p_{att} , the greater the proportion of novel syllables that were rated as unattractive, and thus the stronger the content bias. Finally, conformist biases were included by having individuals sample the repertoires of N_T potential demonstrators, and memorise their syllable types. Then syllables were chosen relative to their frequency within the sample, raised to the power α . Values of α greater than 1 correspond to conformist biases; values less than 1 to anti-conformist biases. These five key parameters (μ , N_T , v , p_{att} , α) each influence the frequency distribution of syllable types (Fig. 2a). The aim of our analysis was to infer values for these parameters that were compatible with the empirical data.

We fit our model to the empirical data-set using a Population Monte Carlo (PMC) variant of ABC³⁷, using a set of 13 summary statistics encompassing different types of information about the frequency distribution and pattern of sharing of syllable types, which we reduced to 6 components using partial least squares^{44,45}. This way, we estimated that swamp sparrows in our recorded populations learn syllables with a mutation rate, μ , of 0.0185 (95% Credible Interval: 0.0062 - 0.0465, see also Supplementary Fig. 3, Supplementary Table 3 for posterior distributions and parameter correlations). We did not find a strong correlation between population size and μ ($r=-0.202$), perhaps due to the role of conformist bias (see Supplementary Note 2). Nevertheless, our credible intervals for mutation rate are likely influenced to some degree by our prior distributions of population size (log-uniform with limits 400 and 3000). This prior was informed by estimates of population size based on our surveys of the populations (Supplementary Table 1). It is harder to rule out extensive migration into and out of the population, and some fraction of the “mutations” in each population may be due to novel songs being introduced in by migrants. Thus our estimate of μ is likely to be higher than the true value.

Estimates of Learning Parameters

We found no evidence to either rule in or out a strong demonstrator bias to learning, with a broad posterior distribution encompassing most of the prior distribution (median estimate of $\nu=1.15$, CI: 0.0146 - 5.70). Our model does exclude the possibility of there being a very strong content bias ($p_{\text{att}} = 0.607$, CI: 0.217 - 0.973). Content bias was negatively correlated with mutation rate ($r=-0.612$): stronger content biases led to many innovated syllable types being removed from the population without being copied by others, and thus allowed a higher mutation rate to be

consistent with the data. The lack of clear evidence for demonstrator and content biases does not mean that they do not exist. In the case of content biases, for example, it is well established that swamp sparrows selectively avoid learning heterospecific song⁴⁶. But it appears that novel songs which fall outside the species-typical range are only generated rarely, at levels at which the bias has only a minute effect on syllable frequency distributions.

We found strong evidence for a conformist bias, with the posterior distribution for parameter α clearly greater than 1 (median: 1.316, CI: 1.079 - 1.700). α was strongly negatively correlated ($r=-0.700$) with the number of potential demonstrators sampled, N_T (median estimate of N_T : 4.71, CI: 2.59 - 23.3). The more demonstrators that individuals sampled during learning, the lower their error in assessing the frequency of syllable types and the weaker the conformist bias necessary to fit the data well. Interestingly, although our evidence for a conformist bias was clear, it was only of moderate strength at an individual level. For example, in the hypothetical situation of a male that sampled 5 demonstrators, each with three syllable types, with $\alpha = 1.316$, if 5 of the syllables were of one type and 10 of them unique, the probability of the male selecting the shared syllable type as the first syllable type in its repertoire would be 0.45 compared to a baseline probability of 0.33 without any conformist bias. At the population level, however, this bias had a large effect on the frequency distribution of syllable types (Fig. 2b) due to the amplifying effects of repeated cultural transmission. In a different context, a theoretical model predicted moderate or weak levels of conformist bias to be favoured when learning precision is high⁴⁷ (as it is in swamp sparrows), and moderate conformist biases were found in great tits that relied more on social learning for a foraging task¹⁹.

Estimates of the stability of syllable-type traditions

Cultural transmission parameters influence the turnover of cultural traits, and thus the stability of traditions over time. Within our simulations, we recorded the date at which each syllable type was innovated, and then measured the age of all syllable types that were present at the end of the simulation run (types that persisted throughout the entire simulation run were assigned an age of 5000 years). The average age of the oldest syllable type in each population was 1537 years, and 8.6% of syllable types (and 26.5% of all syllables) were older than 500 years. More common syllable types tended to be older than younger ones (Fig. 4). Thus our results suggest that the individual song learning behaviour of swamp sparrows is capable of maintaining traditions for extremely long periods of time. Our models cannot take into account large-scale demographic and ecological changes (e.g. the temporary extinction and re-colonisation of a marsh population due to changes in water level), instead these estimates merely suggest that, under stable conditions, syllable type traditions can last for a very long period. Direct anecdotal evidence for the maintenance of swamp sparrow syllable types comes from a visual comparison of syllable types in the Hudson Valley population with spectrograms from an earlier study with recordings made in 1976-8⁴⁸ (31-33 years before our sample). All but two of the 19 syllable types sung by 2 or more males in our 2009 sample were found in the earlier sample, and the commonest type in 2009 was also the commonest type in the 1970's.

An empirically-inspired model of conformist bias

Our evidence for a conformist bias provokes a further question: what is the process underlying that bias? Social learning can be the result of cognitively complex processes (e.g. imitation), but also of simple ones (e.g. local or stimulus enhancement). We argue that the same is true for

conformist biases, and that swamp sparrow conformity may not be based on sophisticated record-keeping of the relative frequencies of different syllable types, but rather on a much simpler process. The process of song overproduction and selective attrition described for swamp sparrow song development²³ provides such a process. In closely related species, it has been demonstrated that during selective attrition, individuals preferentially retain syllable types that they hear others singing^{26,28}. This process of repeatedly comparing memorised syllable types against the vocal output of other individuals generates by itself a conformist bias, because common syllables are more likely to be encountered repeatedly.

We modelled this process as follows: individuals first memorise a sample of N_{T1} adult syllable repertoires. They select their adult repertoire from this memorised sample by removing syllable types, one by one. At each stage, the decision of which type to remove is determined partly by a second random sample of N_{T2} adult repertoires: the more frequently syllable types occur in this second set, the less likely they are to be removed. This process is in line with the idea that birds listen to adults singing around them during this period of selective attrition, and are less likely to remove a syllable type that they have heard recently (note that the lack of geographic structure within populations suggests that this process may often occur before learners acquire their own territory). This process of repeated sampling inherently favours common syllable types over rare ones. We quantified the relative role of the first and second sampling periods during selection with a weight parameter, W . Repeating the fitting process with ABC, we found that this model too was readily able to closely fit the empirical data, with parameter estimates of N_{T1} : 3.08 (CI: 2.53 - 4.42); N_{T2} : 3.6 (CI: 1.6 - 10.2); W : 1.58 (CI: 0.54 - 9.60). N_{T2} was strongly positively correlated with W : the more syllables were sampled during the second phase, the greater the relative role of the first phase. This relationship effectively maintained a constant ratio between

the contributions of the first and second phase of learning to the overall choice of which syllable type to retain in their repertoire. The frequency distribution of syllable types generated by the model under these parameters showed the same relatively high levels of common and rare syllable types as found in both the conformist bias model and the empirical data.

Discussion

We have for the first time been able to derive estimates of the individual behavioural parameters underlying cultural transmission of bird song. Our ABC-fitted models suggest that song learning in swamp sparrows is highly precise. This precise learning is complemented by a conformist bias, which may be generated by the cognitively simple process of overproduction and selective attrition. This hypothesis could be tested with further laboratory experiments investigating how songs heard during the motor phase of song development influence the process of selective attrition and influence the choice of which type is learned. Together, these conditions would allow song traditions to persist for many hundreds of years, and if our model is an accurate reflection of swamp sparrow learning, it would demonstrate that non-human cultural traditions can match the stability of those found in humans.

Our methods can readily be applied to other species of vocal learners for which similar data-sets exist. Among songbirds, it is clear that some species of birds learn precisely, such as swamp sparrows, while others rarely learn all parts of a demonstrator's song precisely⁴⁹. Yet because we have lacked methods to quantify the precision of learning, we have not been able to apply comparative methods to examine the evolutionary causes and consequences of song-learning behaviour. Doing so would allow us, for example, to test hypotheses about the role of sexual

264 selection in the evolution of precise learning⁵⁰ or how learning might influence speciation
265 processes⁵¹.

266 Previous laboratory experiments have demonstrated that swamp sparrows are capable of
267 imitating the songs of tutors with high precision^{22,23,46}. Our results suggest that nearly all songs
268 are learned with such precision, which in turn indicates that swamp sparrows learn syllables as
269 precisely as possible, and avoid singing unique or rare syllable types. This is consistent with
270 experimental evidence that more typical variants of common syllable types are more attractive to
271 female swamp sparrows, and stimulate a more aggressive response from territorial males⁵². The
272 consequences of song learning precision serving as an assessment signal⁵²⁻⁵⁴ may have been
273 selection for precise transmission but also for a simple mechanism of conformist transmission,
274 and, in turn, stable traditions. In human culture, precise transmission and stable traditions have
275 been argued to be a prerequisite for complex, cumulative culture^{1,3,4}. In swamp sparrow song,
276 studies of sub-syllabic note type variation have revealed overlapping and context-dependent
277 patterns of categorical perception that suggest an unusual degree of sophistication and
278 hierarchical structure in cultural evolution⁵⁵⁻⁵⁷. But despite possessing precise transmission that
279 leads to stable traditions, the complexity of swamp sparrow vocal culture of course does not
280 begin to approach that of humans. Our findings, then, suggest that the ability to transmit
281 traditions with precision can no longer be considered a fundamental difference between how
282 human and non-human cultures evolve.

283 Methods

284

285 Song Recordings

286 We recorded the song repertoires of 615 adult male swamp sparrows between May 5th 2008 and
287 July 12th 2009 from six populations across north-eastern USA (Supplementary Table 1). All
288 populations are within the species range of the subspecies *M. g. georgiana*. We recorded the
289 location of each individual's territory using a hand-held GPS device. Recordings were made with
290 a Sony PCM D50 digital recorder at a sampling rate of 44.1kHz and at 16bits, using a Shure
291 SM57 microphone mounted in a Sony PBR 330 parabola. To maximize the chances of recording
292 each of the syllable types in a male's repertoire, we recorded each male for either at least one
293 hour or until it cycled through its supposed repertoire at least 1.5 times^{57,58}. We selected one
294 exemplar for each syllable type in each individual's repertoire, based on a subjective assessment
295 of recording quality.

296

297 Song Measurements

298 Exemplars were measured using the Luscinia sound analysis program
299 (<http://github.com/rflachlan/luscinia>, version 2.17.11.22.01). Spectrograms were produced after
300 high-pass filtering the recording at 1kHz. We standardized spectrogram settings for our
301 recordings. The window length was 5ms, and the time step was 0.5ms. A Hamming window
302 function was used. Recordings were further processed using the echo removal function of
303 Luscinia with an 'echo range' of 25ms and 'echo removal' of between 25 and 100%, checking

visually that the signal was not attenuated by the algorithm. We set the dynamic range to -40dB, relative to the maximum amplitude within the entire spectrogram and set the dynamic equalization function to 200ms to increase the amplitude of the quieter the initial and final syllables within swamp sparrow songs. Using Luscinia, we then measured each element, and marked each syllable within each exemplar.

Syllable Comparison

We compared the structure of each syllable with every other syllable using the dynamic time warping (DTW) algorithm in Luscinia (see ^{52,59} for more details), setting the Compression Factor parameter to 0.001, with a Minimum Element length of 10. This ensured that every element in the data set was characterized by 10 data-points, evenly spread through its length. We used time, fundamental frequency, fundamental frequency change, and vibrato amplitude as the basis for the DTW comparisons (with relative weightings 10, 1, 1, and 0.25). Fundamental frequency was log-transformed.

The outcome of the Luscinia comparison is a matrix of syllable dissimilarities. We clustered syllables on the basis of this matrix using the hierarchical UPGMA algorithm, and assigned syllables to types on the basis of the resulting dendrogram. We calculated a measure of clustering tendency, the global silhouette index (GSI) ⁴³ at each depth within the dendrogram, and searched for peaks in the index, which identify natural clusters among the syllables. This clustering procedure produced a dendrogram with a peak in the GSI at $k = 160$ syllable type clusters (Supplementary Fig. 2). These clusters closely matched labels given to syllables by subjective visual assessment of spectrograms for the Conneaut population (73 human vs 64 DTW clusters;

325 Normalized Mutual Information index of cluster similarity⁶⁰=0.868). We used this clustering
326 solution for all further analyses.

327

328 Analysis of Geographic Variation

329 In order to verify our assumption that any particular adult within a population would be equally
330 likely to be the demonstrator for any particular juvenile, we first tested whether, within
331 populations, syllables tended to be shared with territorial neighbours or not. Using the $k=160$
332 clustering mentioned above, we measured the Jaccard coefficient of repertoire sharing between
333 each pair of individuals in the population, and binned comparisons based on the geographic
334 distance between them. We also carried out Mantel tests of correlation between matrixes of
335 Jaccard coefficients and geographical distance.

336

337 Cultural Evolutionary Models of Song Learning and Conformity

338 We examined two different models of song learning, both of which included a conformist bias:
339 in model 1, this bias was defined explicitly; in model 2 it arose from the process of selective
340 attrition. We define a conformist bias as any tendency to copy common traits at an even greater
341 probability than their frequency in the population. In most cases, however, it is impractical for
342 individuals to assess trait frequencies in an entire population. A more realistic proposition is that
343 individuals sample a subset of the population, and select more frequent traits from that sample.
344 Simulations were initiated with syllable types randomly allocated to individuals, and lasted 5000
345 years, at the end of which summary statistics were calculated.

346 Model 1

347 We constructed an individual-based simulation model of song learning and cultural evolution.
348 Each population consisted of N_p individuals. Each individual in the model learned song syllables
349 during its first year of life only. Syllables were characterized by an integer, x , (range: $1-N_s$).
350 Each syllable type had an intrinsic attractiveness to be learned, M_x , which was determined at the
351 outset of the simulation and remained constant throughout the simulation (with a proportion of
352 syllable types, p_{att} , deemed attractive with $M=1$, and the remainder unattractive with $M=0.05$).
353 Each individual, m , was first assigned a repertoire size r_m by sampling randomly from the
354 distribution of repertoire sizes in our empirical data-set. Then, juvenile individuals filled their
355 repertoire by imitating adults in the population. In each year of the population, individuals had a
356 mortality rate of 0.4 (our results were not sensitive to this parameter). We used years as
357 convenient time-steps for the model since swamp sparrows have one clear breeding season per
358 year, and learn songs only during a fixed period during their first year of life. Individuals that
359 died were immediately replaced. Each male in the population was also assigned at birth a
360 "demonstrator attractiveness index", t_m . This was calculated as: $t_m = e^{\varphi(0, v)}$, where v is a
361 parameter that determined the variability in demonstrator attractiveness in the population.

362 The first step of imitation was to sample a set of N_T potential adult demonstrators from the
363 population and memorise each syllable in their repertoires. If a given syllable type had already
364 been memorised, then it was not added to the memory. But the frequencies of each syllable type
365 in the sample of demonstrators, F_x , was recorded. By raising F_x to a parameter α , conformist
366 ($\alpha>1$) or anti-conformist ($\alpha<1$) biases could be created, where the higher α was, the stronger the
367 bias. Some syllable types were produced by multiple demonstrators. Therefore, in order to model
368 the effects of variation in demonstrator attractiveness, we had to produce an aggregate

demonstrator score for a particular syllable, across all the demonstrators sampled by a particular learner. The simulation thus calculated a combined demonstrator attractiveness score, T_x , for each syllable type, x , in the male's memory, by taking the mean of the demonstrator attractiveness scores, t_m , for each demonstrator that sang that particular type. In the main model, the probability of a male selecting a given syllable type, x , to learn was then set proportional to:

$$Pl_x = F_x^\alpha \cdot M_x \cdot T_x$$

Syllables were learned by sampling from the distribution of Pl without replacement. With a probability μ , individuals produced a different syllable type from that selected ("mutation"). In this case, the syllable type was sampled randomly from the uniform distribution $\{1:N_s\}$.

Model 2

In the second model, learning occurred over two phases. In the first phase, the repertoires of N_{T1} adult demonstrators were memorized, just as in Model 1. In the second phase, the repertoires of a separate sample of N_{T2} adults were memorized. Syllable types that were encountered for the first time in this second phase could not be memorized (i.e. birds had to select from the list of syllable types they memorized in the first phase). The frequency of syllable types encountered in the second phase that had already been memorized in the first phase was given by G . Then, for each syllable type x : $Pr_x = (W + G_x) \cdot M_x \cdot T_x$

Individuals then progressively reduced their repertoire size from that memorized in the first phase: in each step of this process, syllables were selected by sampling without replacement from Pr until one syllable remained unselected, and that syllable was then removed from the repertoire. (If more than 8 syllable types had been memorized in the first phase, the very first reduction step reduced the repertoire to 8. This approach was required to make the simulation

computationally feasible, but it also corresponds to the maximum number of imitated syllable types observed in songs produced by swamp sparrows during the early “plastic song” phase of song development ²³.

Summary Statistics

To compare empirical to simulated data, we measured 13 summary statistics from each population. Inference in ABC is subject to the summary statistics used, and so we used an exhaustive approach, employing as many statistics as possible. These were: (1) Singletons: proportion of syllable types that were sung by only one individual in the sample; (2) Rare syllable types: proportion of syllable types that were sung by 4 individuals or less; (3) Intermediate-frequency syllable types: proportion of syllable types that were sung by more than one but less than or equal to c individuals, where the threshold c was 5% of the total number of syllable types in the sample; (4) Common-frequency syllable types: proportion of syllable types that were sung by more than c individuals; (5) ns , The number of syllable types in the sample; (6) The number of individuals singing the most common syllable type in the sample; (7) H , an index of diversity, calculated as:

$$H = \sum_i^{ns} f_i \cdot \log_2(f_i)$$

where f_i is the proportion of syllables in the sample of type i . (8) α_p , the exponent of the fitted power law function. We fitted a power-law function to the ordered empirical probability distribution of syllable types. α_p was calculated⁶¹ as:

$$\alpha_p = 1 + ns \cdot (\sum_i^{ns} \ln(2 \cdot f_i))^{-1}$$

(9) The degree of fit of the cumulative empirical probability distribution to the fitted power law distribution; specifically the Kolmogorov-Smirnov statistic (the largest absolute discrepancy between the empirical and power-law distribution). (10) A second statistic capturing the fit of the data to a power-law distribution: the absolute discrepancy between the two when the frequency of the syllable type is 2. This captured the fit of the power law distribution at one extreme of the distribution, where we had observed key departures in exploratory work. (11) The proportion of pairs of individuals that shared at least one syllable type. (12) The proportion of pairs of individuals that shared more than one syllable type. (13) The mean correlation-coefficient between syllable types:

$$\bar{r} = \frac{\sum_i^{ns} \sum_j^i r_{i,j} \sqrt{p_i \cdot p_j}}{\sum_i^{ns} \sum_j^i \sqrt{p_i \cdot p_j}}, \text{ where } r_{i,j} = \frac{p_{i,j} - p_i \cdot p_j}{\sqrt{p_i(1-p_j)p_j(1-p_i)}}$$

where $p_{i,j}$ is the probability of types i and j occurring in one bird's repertoire, and p_i is the probability of type i in the sample. Statistics (12) and (13) both capture information relating to the probability that individuals share more than just one syllable type in their repertoires, and are thus intended to be related to a tendency to learn songs from one particular demonstrator. Statistics (1-4, 11, 12) were logit-transformed.

To address potential redundancy among the 13 statistics, and to find the most informative combination of statistics, we used a partial least squares method (PLS)⁴⁴. First we simulated data, drawing 2400 sets of parameter values from priors (see below) and running a simulation for each set. For each simulation run, we measured the 13 summary statistics described above. We then ran a partial least squares analysis, using the simulation parameters as dependent variables and the summary statistics as independent variables,

using the package `pls`⁴⁵. A leave-one-out cross validation suggested that 6 components were best able to explain variation in the simulation parameters on the basis of the summary statistics (Supplementary Fig. 4, 5). The loading weights for these components are given in Supplementary Table 2.

Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) attempts to provide an unbiased estimate of the posterior distribution of parameters ($p(\theta|y)$) by comparing summary statistics taken from empirical data (y) with summary statistics taken from simulations of the empirical data (y_s): $p(\theta|y) \propto Pr(|y_s - y| < \varepsilon | \theta) \cdot p(\theta)$. Simulations are carried out with parameters drawn from prior distributions ($p(\theta)$), and if the summary statistics calculated from the simulations are within a threshold ε of the empirical summary statistics, the parameter values are accepted. A distribution of such accepted parameter values constitute an unbiased estimate of the posterior distribution in the above equation, as $\varepsilon \rightarrow 0$. (see ³⁷ for a general review). In each run of our simulations, we carried out independent simulations for each of the 6 populations. The overall dissimilarity between the simulated and empirical PLS dimensions was calculated as the Euclidean distance, over all PLS dimensions, between the mean of the simulated PLS scores and the mean of the empirical PLS scores.

Since the basic ABC approach can require a very large number of simulations, and since our agent-based simulations were relatively computationally taxing, we employed the Population Monte-Carlo variant of ABC (PMC-ABC, ³⁷). PMC-ABC increases the efficiency of ABC by carrying out a series of sets of simulations in which the threshold ε is progressively reduced. We used the series $\{20, 16, 13, 11, 9, 8, 7, 6, 5, 4, 3, 2\}$. In each set of simulations, we obtained 1000

accepted parameter sets. Because model comparison in ABC is problematic⁶², we adopted a strategy that allowed us to test alternative hypotheses within models.

Prior Distributions

We set the prior distributions for N_p , N_T , N_s , v , α , and μ as log-uniform distribution with limits $N_p: \{400, 3000\}$; $N_s: \{180, 500\}$; $N_T: \{2.5, 50\}$; $v: \{0.01, 6\}$; $\alpha: \{0.25, 4\}$; $\mu: \{0.0001, 0.3\}$. p_{att} was given a uniform distribution $\{0.01, 1\}$. In model 2, N_{T1} , N_{T2} , and W were also given log-uniform distributions with limits $N_{T1}: \{2.5, 15\}$; $N_{T2}: \{1.5, 50\}$; $W: \{0.01, 20\}$. These choices were made to encompass all potential outcomes (e.g. from very strong Model biases to very weak ones) and to be uninformative.

Validation of ABC

We carried out a leave-one-out cross validation analysis to check that our ABC design was providing unbiased parameter estimates. We simulated 100,000 samples directly from the priors, and, taking each sample in turn, estimated parameter values using the remaining samples. Because number of demonstrators sampled, N_T , was strongly correlated with conformist bias, we fixed it at 5. This analysis showed that the two key parameters, conformist bias and mutation rate, were accurately estimated by the analysis (Supplementary Fig. 6). Model bias was not accurately estimated by the analysis, while content bias was only estimated precisely when it was strong ($p_{att} \sim 0.25$). We note that these two latter parameters were included in our model only because of their potential confounding effects on the parameters of primary interest.

473 Data Availability

474 The Luscinia database of measured swamp sparrow songs used in the recording can be found:
475 <https://figshare.com/s/5cb1b37e9bbc9bb00955> and includes all recordings used in the analysis;
476 metadata (including locations of individuals); and song measurements. The fileset also includes
477 the output of the song comparison analysis, and the output of the ABC analysis.

478 Code Availability

479 The simulation code is freely available at <https://github.com/rflachlan/SongABCJPPF>.

480

481 **References**

- 482 1. Boyd, R. & Richerson, P. J. Why culture is common, but cultural evolution is rare. *Proc.*
483 *Brit. Acad.* **88**, 77–94 (1996).
- 484 2. Laland, K. N. & Janik, V. M. The animal cultures debate. *Trends Ecol. Evol.* **21**, 542–547
485 (2006).
- 486 3. Galef, B. G. The question of animal culture. *Hum. Nat.* **3**, 157–178 (1992).
- 487 4. Tomasello, M., Kruger, A. C. & Ratner, H. H. Cultural learning. *Behav. Brain. Sci.* **16**, 495–
488 552 (1993).
- 489 5. Tennie, C., Call, J. & Tomasello, M. Ratcheting up the ratchet: on the evolution of
490 cumulative culture. *Phil. Trans. Roy. Soc. B.* **364**, 2405–2415 (2009).
- 491 6. Claidiere, N. & Sperber, D. Imitation explains the propagation, not the stability of animal
492 culture. *Proc. Roy. Soc. Lond. B* **277**, 651–659 (2010).
- 493 7. Henrich, J. & Boyd, R. The evolution of conformist transmission and the emergence of
494 between-group differences. *Evol. Hum. Behav.* **19**, 215–241 (1998).
- 495 8. Henrich, J. Cultural transmission and the diffusion of innovations: Adoption dynamics
496 indicate that biased cultural transmission is the predominate force in behavioral change.
497 *American Anthropologist* **103**, 992–1013 (2001).
- 498 9. Kandler, A. & Laland, K. N. An investigation of the relationship between innovation and
499 cultural diversity. *Theor. Pop. Biol.* **76**, 59–67 (2009).
- 500 10. Morgan, T. J. H. & Laland, K. N. The biological bases of conformity. *Front Neurosci* **6**, 87
501 (2012).
- 502 11. Whiten, A., Horner, V. & de Waal, F. B. M. Conformity to cultural norms of tool use in
503 chimpanzees. *Nature* **437**, 737–740 (2005).
- 504 12. Dindo, M., Whiten, A. & de Waal, F. B. M. In-group conformity sustains different
505 foraging traditions in capuchin monkeys (*Cebus apella*). *PLoS ONE* **4**, e7858 (2009).
- 506 13. Haun, D. B. M., Rekers, Y. & Tomasello, M. Majority-biased transmission in chimpanzees
507 and human children, but not orangutans. *Current Biology* **22**, 727–731 (2012).
- 508 14. van de Waal, E., Borgeaud, C. & Whiten, A. Potent Social Learning and Conformity Shape
509 a Wild Primate's Foraging Decisions. *Science* **340**, 483–485 (2013).
- 510 15. Cantor, M. *et al.* Multilevel animal societies can emerge from cultural transmission. *Nat.*
511 *Commun.* **6**, 8091 (2015).
- 512 16. Whiten, A. & van de Waal, E. Identifying and dissecting conformity in animals in the wild:
513 further analysis of primate data. *Anim. Behav.* **122**, e1–e4 (2016).
- 514 17. Pike, T. W. & Laland, K. N. Conformist learning in nine-spined sticklebacks' foraging
515 decisions. *Biol. Letters* **6**, 466–468 (2010).
- 516 18. Aplin, L. M. *et al.* Experimentally induced innovations lead to persistent culture via
517 conformity in wild birds. *Nature* **518**, 538–541 (2016).

19. Aplin, L. M., Sheldon, B. C. & McElreath, R. Conformity does not perpetuate suboptimal traditions in a wild population of songbirds. *Proc. Natl. Acad. Sci. USA* **114**, 7830–7837 (2017).
20. Catchpole, C. K. & Slater, P. J. B. *Bird song: biological themes and variations*. (Cambridge University Press, 2008).
21. Podos, J., Huber, S. K. & Taft, B. Bird song: the interface of evolution and mechanism. *Annu. Rev. Ecol. Evol. Syst.* (2004). doi:10.2307/annurev.ecolsys.35.021103.30000004
22. Marler, P. & Peters, S. Sparrows learn adult song and more from memory. *Science* **213**, 780–782 (1981).
23. Marler, P. & Peters, S. Developmental overproduction and selective attrition: new processes in the epigenesis of birdsong. *Dev. Psychobiol.* **15**, 369–378 (1982).
24. Marler, P. & Peters, S. Sensitive periods for song acquisition from tape recordings and live tutors in the swamp sparrow, *Melospiza georgiana*. *Ethology* **77**, 76–84 (1988).
25. Peters, S. & Nowicki, S. Overproduction and attrition: the fates of songs memorized during song learning in songbirds. *Anim. Behav.* **124**, 255–261 (2017).
26. Nelson, D. A. Song overproduction and selective attrition lead to song sharing in the field sparrow (*Spizella pusilla*). *Behav. Ecol. Sociobiol.* **30**, 415–424 (1992).
27. Nelson, D. A., Marler, P. & Morton, M. L. Overproduction in song development: an evolutionary correlate with migration. *Anim. Behav.* **51**, 1127–1140 (1996).
28. Nordby, J. C., Campbell, S. E. & Beecher, M. D. Selective attrition and individual song repertoire development in song sparrows. *Anim. Behav.* **74**, 1413–1418 (2007).
29. Acerbi, A., van Leeuwen, E. J. C., Haun, D. B. M. & Tennie, C. Conformity cannot be identified based on population-level signatures. *Sci. Rep.* 1–9 (2016). doi:10.1038/srep36068
30. Slater, P. J. B. & Ince, S. A. Cultural evolution in chaffinch song. *Behaviour* **71**, 146–166 (1979).
31. Nelson, D. A., Hallberg, K. & Soha, J. A. Cultural evolution of puget sound white-crowned sparrow song dialects. *Ethology* **110**, 879–908 (2004).
32. Derryberry, E. P. Evolution of bird song affects signal efficacy: an experimental test using historical and current signals. *Evolution* **61**, 1938–1945 (2007).
33. Pipek, P. *et al.* Dialects of an invasive songbird are preserved in its invaded but not native source range. *Ecography* **35**, Miskelly–9 (2017).
34. Podos, J. & Warren, P. S. The evolution of geographic variation in birdsong. *Adv. Study Behav.* **37**, 403–458 (2007).
35. Lynch, A. in *Ecology and Evolution of Acoustic Communication in Birds* (eds. Kroodsma, D. E. & Miller, E. K.) 181–197 (Comstock, 1996).
36. Lachlan, R. F. & Slater, P. J. B. Song learning by chaffinches: how accurate, and from where? *Anim. Behav.* **65**, 957–969 (2003).

37. Beaumont, M. A. Approximate Bayesian Computation in Evolution and Ecology. *Annu. Rev. Ecol. Evol. Syst.* **41**, 379–406 (2010).
38. van der Vaart, E., Beaumont, M. A., Johnston, A. S. A. & Sibly, R. M. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling* **312**, 182–190 (2015).
39. Crema, E. R., Edinborough, K., Kerig, T. & Shennan, S. J. An Approximate Bayesian Computation approach for inferring patterns of cultural evolutionary change. *J. Archaeol. Sci.* **50**, 160–170 (2014).
40. Kandler, A. & Shennan, S. A generative inference framework for analysing patterns of cultural change in sparse population data with evidence for fashion trends in LBK culture. *J. Royal Soc. Interface* **12**, 20150905–12 (2015).
41. Crema, E. R., Kandler, A. & Shennan, S. Revealing patterns of cultural transmission from frequency data: equilibrium and non-equilibrium assumptions. *Sci. Rep.* 1–10 (2017). doi:10.1038/srep39122
42. Rendell, L. *et al.* Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends. Cog. Sci.* **15**, 68–76 (2011).
43. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
44. Wegmann, D., Leuenberger, C. & Excoffier, L. Efficient Approximate Bayesian Computation Coupled With Markov Chain Monte Carlo Without Likelihood. *Genetics* **182**, 1207–1218 (2009).
45. Mevik, B.-H. & Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Softw.* **18**, 1–23 (2007).
46. Marler, P. & Peters, S. Selective vocal learning in a sparrow. *Science* **198**, 519–521 (1977).
47. Kandler, A. & Laland, K. N. Tradeoffs between the strength of conformity and number of conformists in variable environments. *J. Theor. Biol.* **332**, 191–202 (2013).
48. Marler, P. & Pickert, R. Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Anim. Behav.* **32**, 673–689 (1984).
49. Mann, N. I. & Slater, P. J. B. Song tutor choice by zebra finches in aviaries. *Anim. Behav.* **49**, 811–820 (1995).
50. Nowicki, S., Peters, S. & Podos, J. Song learning, early nutrition and sexual selection in songbirds. *Am. Zool.* **38**, 179–190 (1998).
51. Lachlan, R. F. & Servedio, M. R. Song learning accelerates allopatric speciation. *Evolution* **58**, 2049–2063 (2004).
52. Lachlan, R. F., Anderson, R. C., Peters, S., Searcy, W. A. & Nowicki, S. Typical versions of learned swamp sparrow song types are more effective signals than are less typical versions. *Proc. Roy. Soc. Lond. B* **281**, 20140252–20140252 (2014).
53. Searcy, W. A., Peters, S., Kipper, S. & Nowicki, S. Female response to song reflects male developmental history in swamp sparrows. *Behav. Ecol. Sociobiol.* **64**, 1343–1349 (2010).

54. Lachlan, R. F. & Nowicki, S. How Reliable Is Song Learning Accuracy as a Signal of Male Early Condition? *Am. Nat.* **180**, 751–761 (2012).
55. Nelson, D. A. & Marler, P. Categorical perception of a natural stimulus continuum: birdsong. *Science* **244**, 976–978 (1989).
56. Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S. & Mooney, R. Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* **12**, 221–228 (2009).
57. Lachlan, R. F. & Nowicki, S. Context-dependent categorical perception in a songbird. *Proc. Natl. Acad. Sci. USA* **112**, 1892–1897 (2015).
58. Ballentine, B. Morphological adaptation influences the evolution of a mating signal. *Evolution* **60**, 1936–1944 (2006).
59. Lachlan, R. F. *et al.* The progressive loss of syntactical structure in bird song along an island colonization chain. *Current Biology* **23**, 1896–1901 (2013).
60. Zhong, S. & Ghosh, J. Generative model-based document clustering: a comparative study. *Knowl Inf Syst* **8**, 374–384 (2005).
61. Clauset, A., Shalizi, C. R. & Newman, M. Power-law distributions in empirical data. *SIAM review* (2009).
62. Robert, C. P., Cornuet, J.-M., Marin, J.-M. & Pillai, N. S. Lack of confidence in approximate Bayesian computation model choice. *Proc. Natl. Acad. Sci. USA* **108**, 15112–15117 (2011).

Acknowledgments

Funding was provided by the Duke University Provost Office and U.S. National Science Foundation grant IOS-1144991 to SN; and the Bill & Melinda Gates Foundation OPP1084362 to OR. Permission and assistance with fieldwork was kindly provided by Montezuma National Wildlife Refuge, Horicon NWR, Thompson Ponds, Phyllis Haehnle Memorial Sanctuary, NY Department of Environmental Conservation, Hudson River National Estuarine Research Reserve, the Pennsylvania State Game Commission and the Pymatuning Laboratory of Ecology.

Author Contributions

RFL conceived the study, recorded and measured the songs, designed and implemented the simulations and analyses, and wrote the paper. OR contributed to the design of the ABC analysis and the writing of the paper. SN contributed to the design of the study and to the writing of the paper.

The authors declare no competing interests, financial or otherwise.

Correspondence and requests for materials should be sent to RFL: r.f.lachlan@qmul.ac.uk

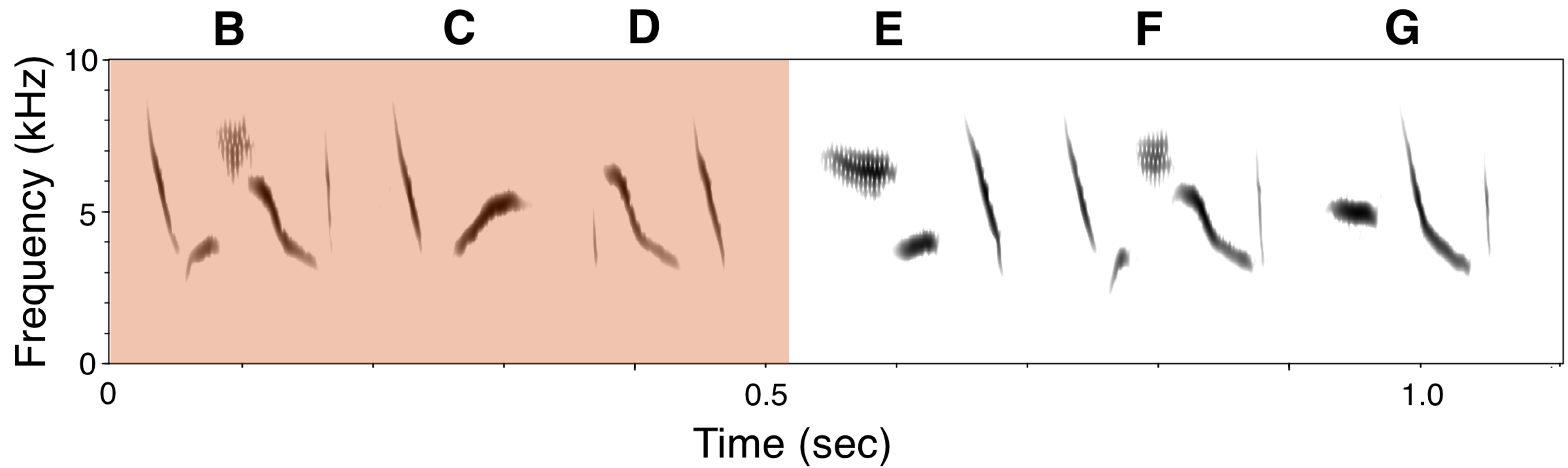
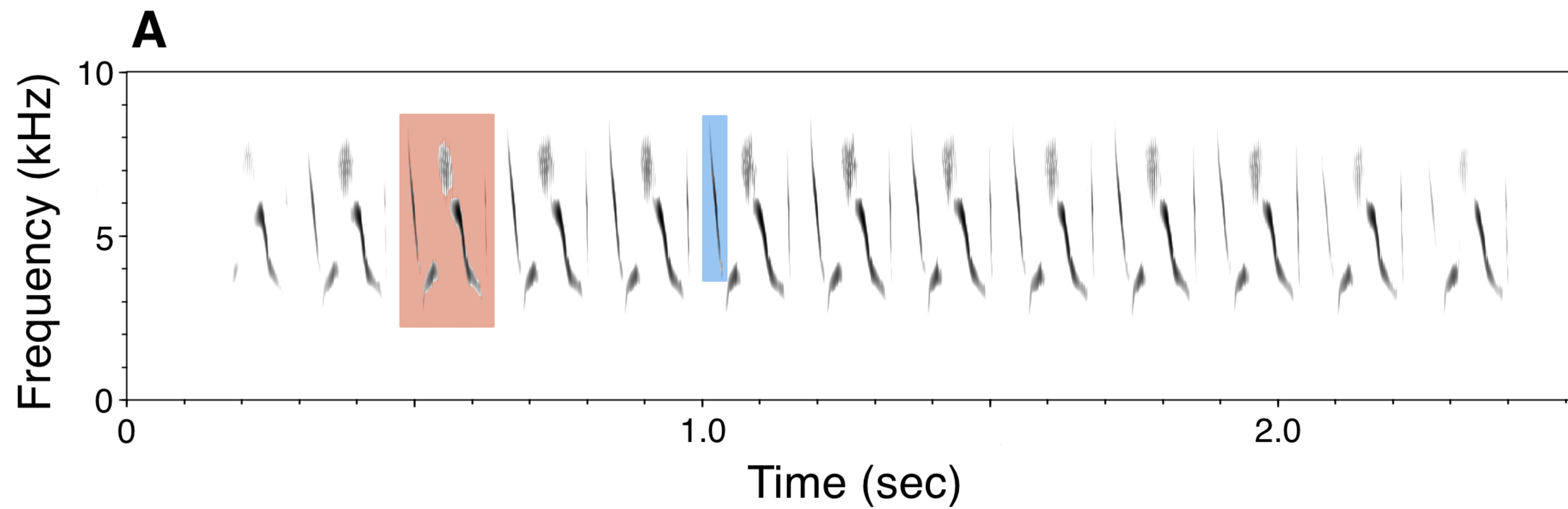
639 Figure Legends

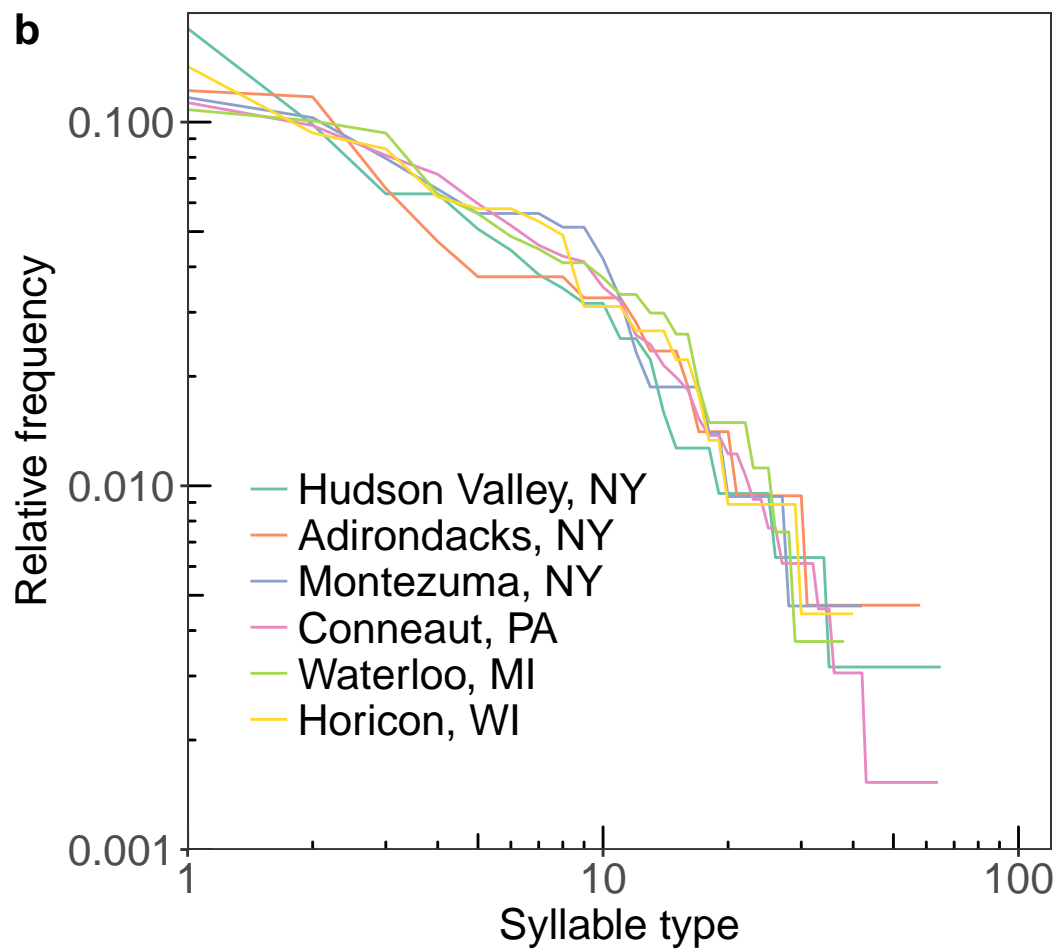
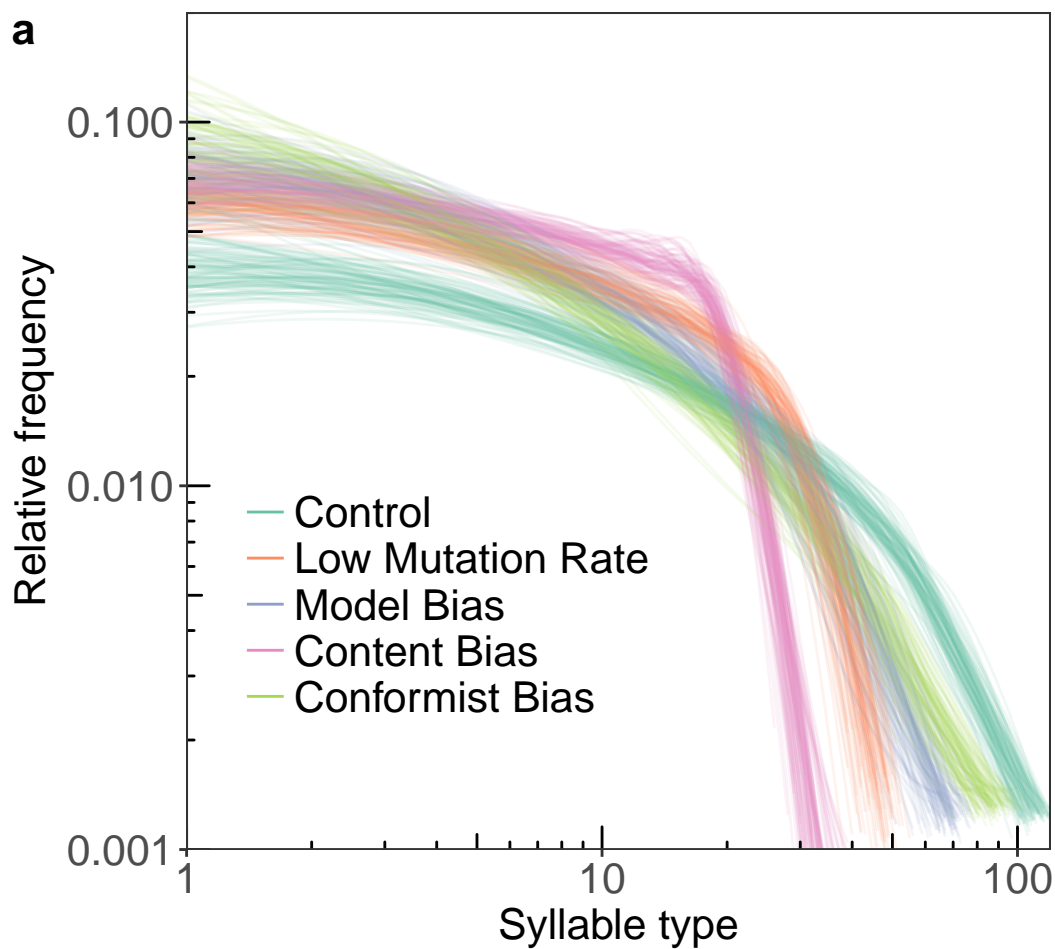
640 **Fig. 1 A** Spectrogram of a swamp sparrow song sung by an individual from the Hudson Valley
641 population. Swamp sparrow songs consist of one syllable (red) repeated 10 or more times. Each
642 syllable consists of 2-5 elements or notes. **B-G** examples of different syllable types. **B-D** show
643 the syllables that make up the repertoire of the same individual whose song is shown in **A**. **E-G**
644 show three syllable types sung by other males in the population, illustrating both the
645 considerable diversity in syllable structure found within a population, but also how, as a
646 consequence of vocal learning, different individuals also sometimes share the same song-type (**B**,
647 **F**).

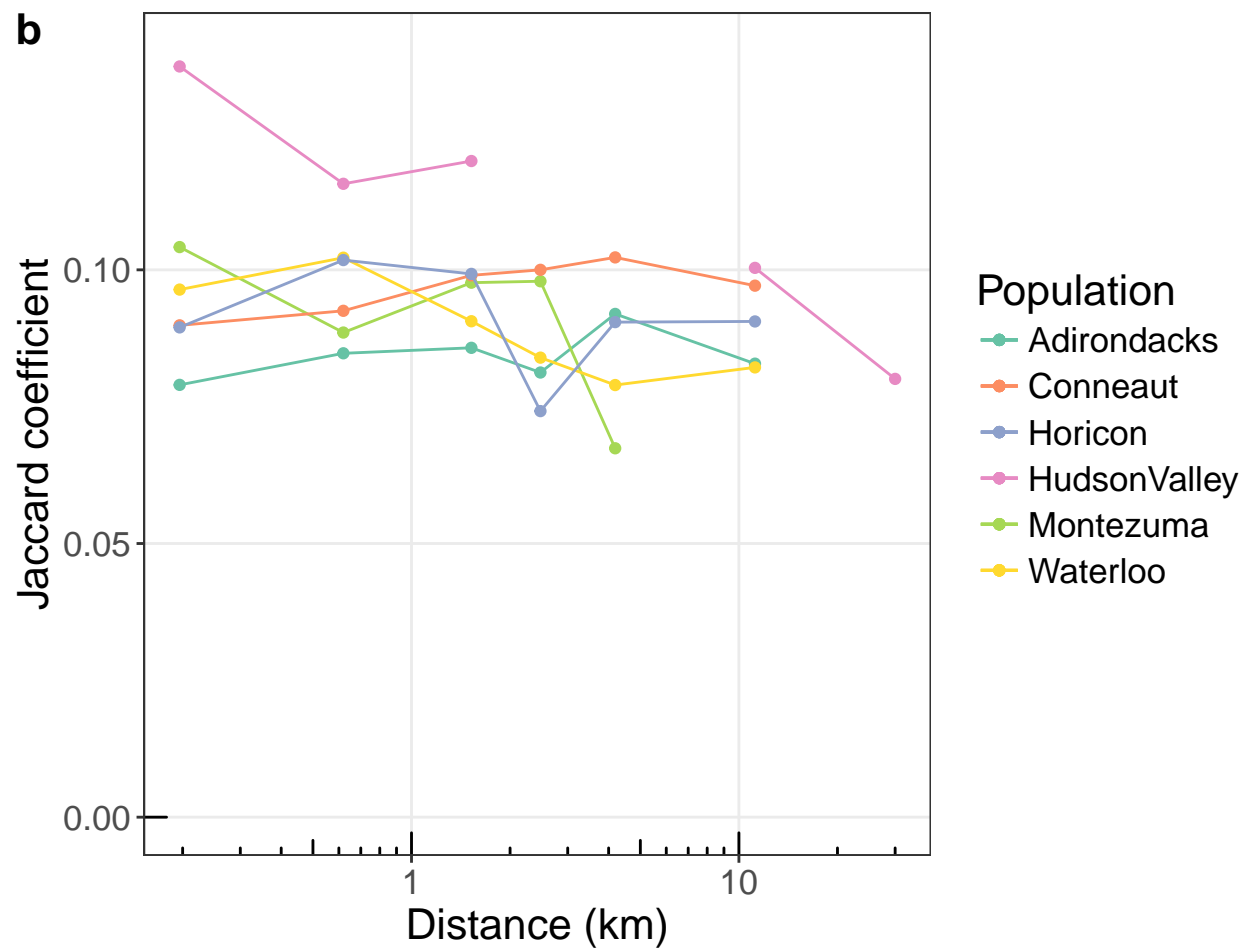
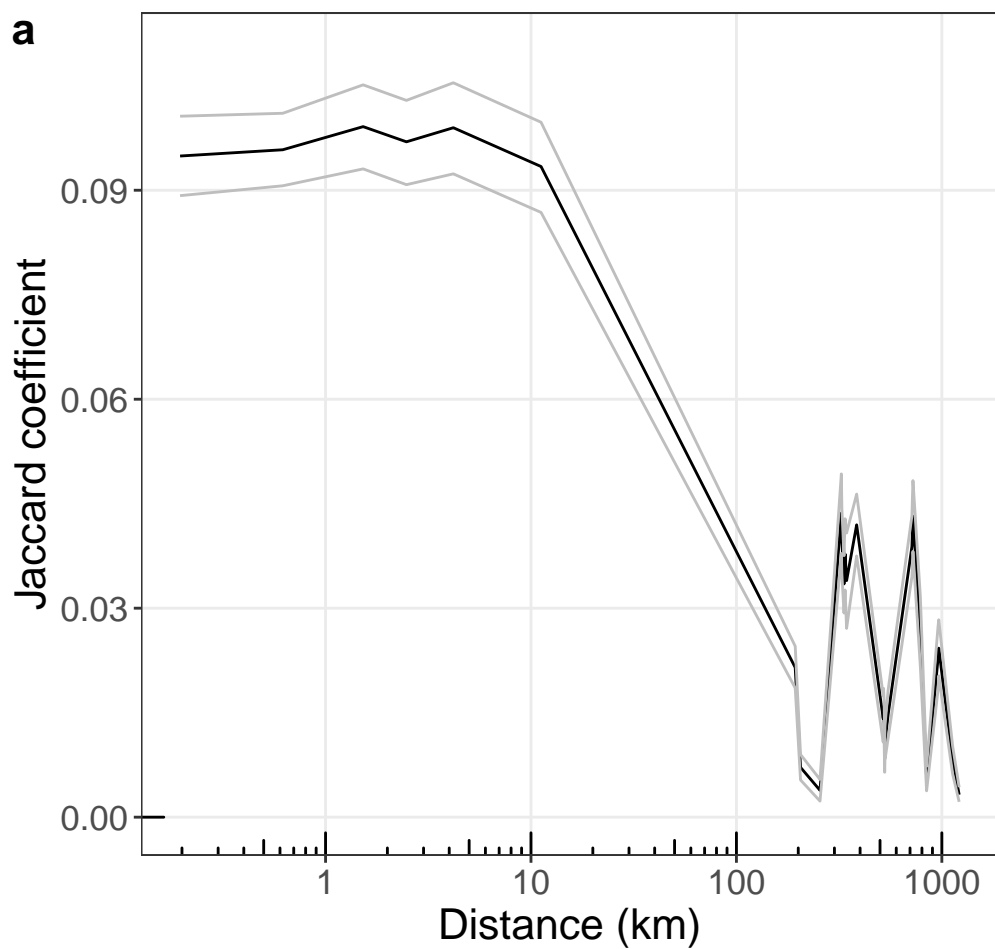
648 **Fig. 2 A** Frequency distributions of syllable types found in simulations with different types of
649 bias, illustrating how different aspects of individual learning influence population-levels of
650 diversity. 100 simulations were carried out for each of 6 conditions. In the Control case,
651 individuals had a mutation rate of 0.01 and no bias to learning. In the Low-mutation rate
652 condition, $\mu=0.001$. In the Model Bias simulations, $v=2$. In the Content Bias simulations,
653 $p_{att}=0.1$. In the Conformist Bias simulations, $\alpha=1.2$. **B** Frequency distributions of syllable types
654 found in the empirical data-set, from the 6 populations sampled (Conneaut, $n=208$, Montezuma,
655 $n=71$, Adirondacks, $n=70$, Waterloo, $n=74$, Horicon, $n=91$, Hudson Valley, $n=101$), illustrating
656 the consistency found between populations and suggesting a match between the empirical data
657 and the simulations with Conformist Bias.

Fig. 3 Geographic variation in syllable sharing. **A** and **B** show the amount of syllable sharing between individuals' (n=615) repertoires, measured using the Jaccard Index (proportion of shared syllable types out of all types found in a pair of individuals), against the geographical distance between them. **A** Syllable sharing was high within populations, with no significant within-population geographic structure, but much lower between populations. **B** Syllable sharing was similar between the six populations sampled.

Fig. 4 Age of syllable types in fitted simulations. The age of syllable types in the 1000 simulations that best fit the empirical data. Each point represents a combination of age in years since the syllable type was innovated and frequency within the population (point size represents the number of times that combination arose). More common syllable types tended to be older, and many syllable types are estimated to be older than 500 years.







Age (years)

1000

10

0.01

0.10

Song-type relative frequency

Count

1000

2000

3000

4000

